

DSM TN 23. Fish density estimation in a zero inflated field with doubly truncated geometric sampling

Ken B. Newman and Lara Mitchell, January 5, 2017

This note builds on TN 18 which described how to estimate fish density when doubly truncated geometric sampling is carried out. Doubly truncated geometric sampling means that a minimum number of samples (tows) are made, and sampling proceeds until at least one fish is caught or an upper bound on the number of samples is reached. There are two processes at work: the mechanism that generates the distribution of the fish and the sampling process at randomly chosen locations and subsequent number of fish caught. In TN 18 the assumption was that the fish density was a homogeneous Poisson process and the catch in a given tow was then $\text{Poisson}(\delta v)$ where δ was the density and v the volume sampled. The robustness of estimators formulated in TN 18 was evaluated for a heterogeneous density field and zero-inflated heterogeneous density field. The best estimator (biased adjusted ratio of means estimator) performed equally well for homogeneous and heterogeneous density fields, but it, and other estimators, performed poorly and inconsistently for zero-inflated fields.

The purpose of this note is to develop estimates of the average density for a zero inflated density field when data are collected according to the doubling truncated geometric sampling distribution. Three different case are considered and presented in order of increasing complexity, summarized below:

1. Zero inflated Poisson with constant tow volumes (ZIP- \bar{v}).
2. Zero inflated Poisson with varying tow volumes (ZIP- v_1, v_2, \dots).
3. Zero inflated Negative Binomial with varying tow volumes (ZINB- v_1, v_2, \dots).

1 ZIP- \bar{v}

For a randomly selected point in the field it is assumed that the probability that the density is zero is denoted π_0 . This would be the case if the fractional area of the region covered by such patches with no fish present was π_0 . Densities in the remainder of the region (the fraction $1-\pi_0$) are assumed a constant, δ , and the catch in a single tow of volume \bar{v} is assumed $\text{Poisson}(\bar{v}\delta)$. Catch in a single tow at a randomly chosen location is then a zero inflated Poisson, ZIP($\pi_0, \bar{v}\delta$), random variable. The likelihood equation for estimating π_0 and δ based upon catch data collected from randomly selected locations where doubly truncated geometric sampling was conducted at each sample location is derived in the following.

1.1 Probability distribution for tows and catches

We begin with the probability distribution for the number of tows, q , in a zero inflated density field and note that the derivation follows easily from TN 18. The number of tows ranges from $q=2$ to $q=q_m$, and if the density is zero at a location, then q is necessarily q_m , while if q is

anything less than q_m , then the density could not be zero. This reasoning leads to marginal distribution for q shown in Table 1.

The distribution of catches conditional on tows is shown in Table 2. When $q < q_m$, catch is a zero-truncated Poisson random variable (as in TN 18). When $q = q_m$ and $C = 0$, then the density could be either zero or positive. The joint distribution of tows and catches is shown in Table 3 and the marginal distribution for catches is shown in Table 4.

Table 1: Marginal probability distribution for the number of a tows q (a doubly truncated geometric random variable) in a zero inflated density field, where the non-zero densities are a constant δ , and tow volume is a constant v .

q	$\Pr(q)$
2	$(1 - \exp(-2v\delta))(1 - \pi_0)$
3 to $q_m - 1$	$\exp(-(q-1)v\delta)(1 - \exp(-v\delta))(1 - \pi_0)$
q_m	$\exp(-(q_m-1)v\delta)(1 - \pi_0) + \pi_0$

Table 2: Conditional probability distribution for catch (C) given the number of tows (q) with doubly truncated geometric sampling in a zero inflated density field, where the non-zero densities are a constant δ , and tow volume is v .

q	$C q$	$\Pr(C q)$
2	$C = 1, 2, \dots$	$\frac{\exp(-2v\delta)(2v\delta)^C}{(1-\exp(-2v\delta))C!}$
$3 \leq q \leq q_{m-1}$	$C = 1, 2, \dots$	$\frac{\exp(-v\delta)(v\delta)^C}{(1-\exp(-v\delta))C!}$
q_m	$C = 0$	$\exp(-v\delta)(1 - \pi_0) + \pi_0$
	$C = 1, 2, \dots$	$\frac{\exp(-v\delta)(v\delta)^C}{C!}(1 - \pi_0)$

Table 3: Joint probability distribution for tows (q) and catch (C) with doubly truncated geometric sampling in a zero inflated density field, where the non-zero densities are a constant δ , and tow volume is v .

q, C	$\Pr(q, C)$
$q = 2, C = 1, 2, \dots$	$\frac{\exp(-2v\delta)(2v\delta)^C}{C!}(1 - \pi_0)$
$3 \leq q \leq q_{m-1}, C = 1, 2, \dots$	$\frac{\exp(-v\delta)(v\delta)^C}{C!}(1 - \pi_0)$
$q = q_m, C = 0$	$\exp(-vq_m\delta)(1 - \pi_0) + \pi_0$
$q = q_m, C = 1, 2, \dots$	$\frac{\exp(-vq_m\delta)(v\delta)^C}{C!}(1 - \pi_0)$

The expected number and the variance of the number of tows can be derived using Table 1, and results from TN 18. The notation q_{non-ZI} refer to the number of tows in the non-zero inflated case (TN 18), and the equations for $E[q_{non-ZI}]$, $E[q_{non-ZI}^2]$, and $Var[q_{non-ZI}]$ can be found in TN 18. The results:

$$E[q] = E[q_{non-ZI}](1 - \pi_0) + \pi_0 q_m \quad (1)$$

$$Var[q] = E[q_{non-ZI}^2](1 - \pi_0) + q_m^2 \pi_0 - E[q]^2 \quad (2)$$

Table 4: Marginal probability distribution for catch (C) with doubly truncated geometric sampling in a zero inflated density field, where the non-zero densities are a constant δ , and tow volume is v .

C	$\Pr(C)$
0	$\exp(-q_m v \delta)(1 - \pi_0) + \pi_0$
1,2,...	$\left(\frac{\exp(-2v\delta)(2v\delta)^C}{C!} + \frac{\exp(-3v\delta) - \exp(-v\delta(q_m+1))}{1 - \exp(-v\delta)} \frac{(v\delta)^C}{C!} \right) (1 - \pi_0)$

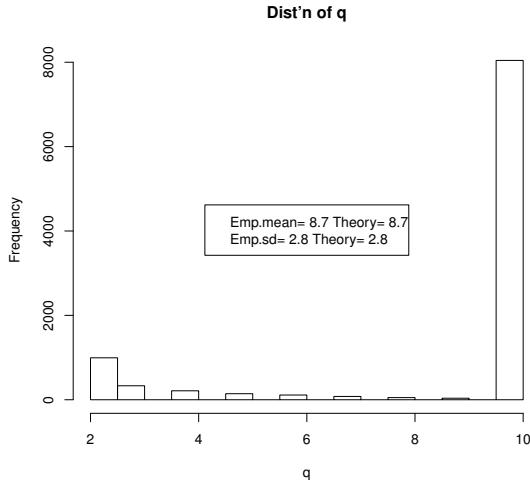
The expected catch is

$$E[C] = E[C_{non-ZI}](1 - \pi_0) = (1 - \pi_0)v\delta \left(2 + \frac{\exp(-2v\delta) - \exp(-q_m v \delta)}{1 - \exp(-v\delta)} \right) \quad (3)$$

where C_{non-ZI} refers to the non-zero inflated case and $E[C_{non-ZI}]$ is shown in TN 18. The variance of the catch has not yet been worked out.

Appendix B contains R code that calculates the distribution for q as well as the mean and variance of q and the expected catch. Figure 1 shows a simulated sampling distribution of q for the case where $\pi_0=0.8$, $\delta=0.5/10,000m^3$, $q_m=10$, and $v=7000m^3$. The probability mass at $q = q_m$ is relatively large given the relatively large π_0 and relatively low δ .

Figure 1: Simulation of the distribution of q , the number of tows, for the doubly truncated geometric sampling procedure in a zero inflated density field, with $\pi_0=0.8$, $\delta=0.5/10,000m^3$, $q_m=10$, and $v=7000m^3$.



1.2 Estimating fish density

Fish density can be estimated simultaneously with π_0 by maximum likelihood using the joint probability distribution of (q, C) shown in Table 3. The log likelihood for a sample of size n is

the following.

$$\begin{aligned}
 l(\pi_0, \delta) \propto & \sum_{i=1}^n I_{q_i=2} [-2v\delta + C_i \ln(2v\delta) + \ln(1 - \pi_0)] + \\
 & I_{3 \leq q_i \leq (q_m-1)} [-q_i v\delta + C_i \ln(v\delta) + \ln(1 - \pi_0)] + \\
 & I_{q=q_m, C_i=0} \ln [\exp(-q_m v\delta)(1 - \pi_0) + \pi_0] + \\
 & I_{q=q_m, C_i>0} [-q_m v\delta + C_i \ln(v\delta) + \ln(1 - \pi_0)]
 \end{aligned} \tag{4}$$

There is no closed form analytic solution for δ and π_0 and numerical methods must be used.

Appendix C.1 has R code for calculating the mles and standard errors.

Estimating abundance. To estimate abundance in a given stratum, the average density for the entire stratum is used, and is $\delta(1 - \pi_0)$. Thus the estimated abundance for stratum h with total volume V_h is

$$\hat{n}_h = V_h \hat{\delta}(1 - \hat{\pi}_0) \tag{5}$$

1.3 Demonstration of mle calculation for a ZIP- \bar{v}

The performance of the maximum likelihood estimates of π_0 and δ was assessed via simulation of tows and catches for different values of π_0 and δ as well as different sample sizes (n) and maximal tow number (q_m).

There were $n=5$ sampling locations and $\pi_0=0.3$, $\delta=0.5/10,000m^3$, $q_m=10$, and $v=7000m^3$. The resulting number of tows and catches were:

q	2	2	4	3	10
C	1	1	1	1	0

The mles were $\hat{\pi}_0 = 0.163$ (SE=0.205) and $\hat{\delta}=0.444/10,000m^3$ (SE=0.36/10000). The contour plot of the negative log likelihood is shown in Figure 2. The true values and mles are shown on the plot along with crude 95% confidence intervals (based on ± 2 SE). The standard errors are relatively large, which is why the crude interval for π_0 is exceeding (0,1), and profile likelihood confidence intervals would be better. The size of the standard errors is a function of the sample size. Figure 3 compares the likelihood surfaces, point and interval estimates for a pair of simulations with $n=5$ versus $n=50$ locations, and the increased precision for $n=50$ is apparent.

Simulations of the performance of the mles for different values of π_0 and δ were also carried out, but the results are similar to simulations for more realistic models, namely varying volumes and heterogeneous densities, which will be shown later (Section 3.3). One thing to emphasize here, however, is that situations where the probability is relatively large that no fish are caught at any sample location result in biases in estimates of π_0 and δ . This makes sense as the prima facie evidence is that $\pi_0=1$. Although the catch of a single fish, which could occur given a larger sample size, would refute that.

Figure 2: Contour plot of log likelihood for a simulation of $n=5$ sampling locations where $\pi_0=0.3$, $\delta=0.5/10,000m^3$, $q_m=10$, $v=7000m^3$. Blue T marks true value and red E marks point estimate.

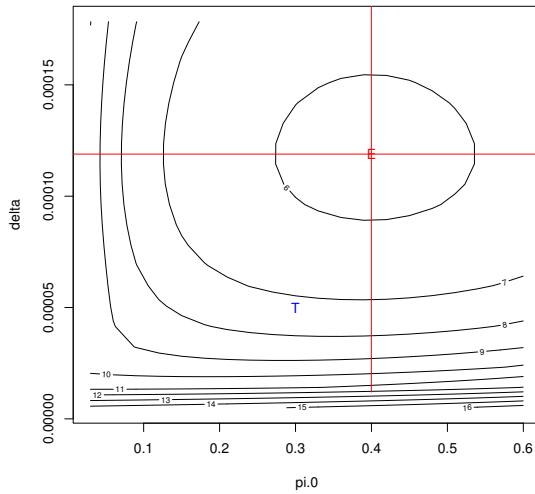
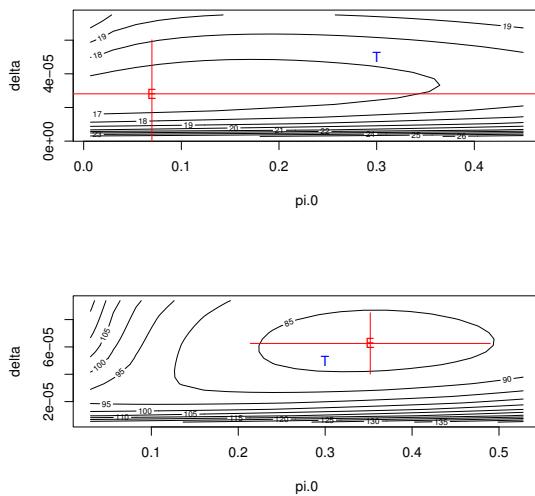


Figure 3: Comparison of contour plots of log likelihood for a simulation of $n=5$ (top plot) and $n=50$ (bottom plot) sampling locations where $\pi_0=0.3$, $\delta=0.5/10,000m^3$, $q_m=10$, $v=7000m^3$. Blue T marks true value and red E marks point estimate.



2 ZIP- (v_1, v_2, \dots)

The probability distributions for tows and catches (marginal, joint, etc) when volume varies between tows (Tables 5 - 8) are minor modifications of the constant volume cases shown previously in Tables 1 to 4.

Table 5: Marginal probability distribution for the number of tows q (a doubly truncated geometric random variable) in a zero inflated density field, where the non-zero densities are a constant δ , and tow volumes (v_i) vary between tows.

q	$\Pr(q)$
2	$1 - \exp(-(v_1 + v_2)\delta)(1 - \pi_0)$
$3 \leq q \leq q_{m-1}$	$\exp(-\sum_{i=1}^{q-1} v_i \delta)(1 - \exp(-v_q \delta))(1 - \pi_0)$
q_m	$\exp(-\sum_{i=1}^{q_m-1} v_i \delta))(1 - \pi_0) + \pi_0$

Table 6: Conditional probability distribution for catch (C) given the number of tows (q) with doubly truncated geometric sampling in a zero inflated density field, where the non-zero densities are a constant δ , and tow volumes (v_i) vary between tows.

q	C	$\Pr(C q)$
2	$C = 1, 2, \dots$	$\frac{\exp(-(v_1+v_2)\delta)((v_1+v_2)\delta)^C}{(1-\exp(-(v_1+v_2)\delta))C!}$
$3 \leq q \leq q_{m-1}$	$C = 1, 2, \dots$	$\frac{\exp(-v_q \delta)(v_q \delta)^C}{(1-\exp(-v_q \delta))C!}$
q_m	$C = 0$	$\exp(-v_{q_m} \delta)(1 - \pi_0) + \pi_0$
	$C = 1, 2, \dots$	$\frac{\exp(-v_{q_m} \delta)(v_{q_m} \delta)^C}{C!}(1 - \pi_0)$

Table 7: Joint probability distribution for tows (q) and catch (C) with doubly truncated geometric sampling in a zero inflated density field, where the non-zero densities are a constant δ , and tow volumes (v_i) vary between tows.

q, C	$\Pr(q, C)$
$q = 2, C = 1, 2, \dots$	$\frac{\exp(-(v_1+v_2)\delta)((v_1+v_2)\delta)^C}{C!}(1 - \pi_0)$
$3 \leq q \leq q_{m-1}, C = 1, 2, \dots$	$\frac{\exp(-\sum_{i=1}^q v_i \delta)(v_q \delta)^C}{C!}(1 - \pi_0)$
$q = q_m, C = 0$	$\exp(-\sum_{i=1}^{q_m} v_i \delta)(1 - \pi_0) + \pi_0$
$q = q_m, C = 1, 2, \dots$	$\frac{\exp(-\sum_{i=1}^{q_m} v_i \delta)(v_{q_m} \delta)^C}{C!}(1 - \pi_0)$

The expected catch in this case of varying volumes is the following:

$$E[C] = (1 - \pi_0)\delta \left((v_1 + v_2) + \sum_{q=3}^{q_m} v_q \exp\left(-\sum_{i=1}^{q-1} v_i \delta\right) \right) \quad (6)$$

Note that this result is the same as (3) when all v_q are a constant.

The log likelihood for the varying volume case, shown below in (7), follows from the joint

Table 8: Marginal probability distribution for catch (C) with doubly truncated geometric sampling in a zero inflated density field, where the non-zero densities are a constant δ , and tow volumes (v_i) vary between tows.

C	$\Pr(C)$
0	$\exp(-\sum_{i=1}^{q_m} v_i \delta)(1 - \pi_0) + \pi_0$
$C = 1, 2, \dots$	$\left(\frac{\exp(-(v_1+v_2)\delta)((v_1+v_2)\delta)^C}{C!} + \sum_{q=3}^{q_m} \frac{\exp(-\sum_{i=1}^q v_i \delta)(v_q \delta)^C}{C!} \right) (1 - \pi_0)$

probability distribution shown in Table 7 and parallels the constant volume case shown in (4).

$$\begin{aligned}
 l(\pi_0, \delta) \propto & \sum_{i=1}^n I_{q_i=2} [-(v_1 + v_2)\delta + C_i \ln((v_1 + v_2)\delta) + \ln(1 - \pi_0)] + \\
 & I_{3 \leq q_i \leq (q_m-1)} \left[-\sum_{j=1}^{q_i} v_{i,j} \delta + C_i \ln(v_{i,q_i} \delta) + \ln(1 - \pi_0) \right] + \\
 & I_{q=q_m, C_i=0} \ln \left[\exp \left(-\sum_{j=1}^{q_m} v_{i,j} \delta \right) (1 - \pi_0) + \pi_0 \right] + \\
 & I_{q=q_m, C_i>0} \left[-\sum_{j=1}^{q_m} v_{i,j} \delta + C_i \ln(v_{i,q_m} \delta) + \ln(1 - \pi_0) \right]
 \end{aligned} \tag{7}$$

Appendix C.2 has R code for calculating the mle for π_0 and δ in the varying volume case.

3 ZINB- (v_1, v_2, \dots)

Two ways were considered for generating overdispersed catches in the non-structural zero portion of the density field: (1) hierarchical modeling of the Poisson rate (density) parameters δ , simulating a heterogenous density field and then draw Poisson samples based on the random densities, and (2) to use a negative binomial distribution for catches.

3.1 Hierarchical modeling of catches: Gamma-Poisson

Heterogeneity in the density field can be simulated from any distribution taking on strictly positive values. Here we used a Gamma distribution for simulating densities at randomly sampled points. The parameterization used was $\text{Gamma}(\alpha, \beta)$, where α is the shape parameter, and β is the scale parameter such that $E[X] = \alpha\beta$ and $Var[X] = \alpha\beta^2$. The pdf is the following:

$$\frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} \exp(-x/\beta).$$

The parameters α and β are calculated as functions of a specified average density value, δ , and coefficient of variation (CV).

$$\begin{aligned}\alpha &= \frac{1}{CV^2} \\ \beta &= CV^2\mu_\delta \\ \delta_i &\sim \text{Gamma}(\alpha, \beta), \quad i = 1, \dots, n \\ C_{ij} &\sim \text{Poisson}(\delta_i v), \quad j = 1, \dots, q_i\end{aligned}$$

3.2 Negative binomial

In the case where the Poisson rate parameter is generated from a Gamma distribution, the distribution for catches is a negative binomial (see Appendix A). Given the equivalency of the approaches, a zero inflated negative binomial model is examined here. The negative binomial model parameterization for the catch in a single tow with volume v , conditional on fish being present, is the following:

$$\Pr(C; v) \sim \text{NegBin}(\alpha\beta v, \alpha) \quad (8)$$

where α and β are identical to those for the hierarchical Gamma-Poisson model.

The likelihood for the variable volume case shown in (7) is extended for this situation. There are now three parameters, π_0 , α , and β , or $\pi_0, \mu_\delta, CV_\delta$. To begin, the marginal distribution for the number of tows, q , is modified by changing the probability of not catching a fish (conditional on fish being present) from the Poisson scenario ($\exp(-\delta v)$) to the negative binomial case:

$$p(0; v) \equiv \Pr(C = 0 | v, \mu_\delta > 0) = \left(\frac{1}{1 + \beta v} \right)^\alpha \quad (9)$$

The resulting distribution for q is shown in Table 9.

Table 9: Marginal probability distribution for the number of tows q (a doubly truncated geometric random variable) in a zero inflated negative binomial density field and tow volumes (v_i) vary between tows.

q	$\Pr(q)$
2	$(1 - p(0; v_1)p(0; v_2))(1 - \pi_0)$
$3 \leq q \leq q_{m-1}$	$\prod_{i=1}^{q-1} p(0; v_i)(1 - p(0; v_q))(1 - \pi_0)$
q_m	$\prod_{i=1}^{q_m-1} p(0; v_i)(1 - \pi_0) + \pi_0$

The conditional distribution for catch given q is shown in Table 10. The probability mass function for the negative binomial, written as $p(C; v)$, is the following.

$$p(C; v) \equiv p(C; \alpha, \beta, v) = \frac{\Gamma(C + \alpha)}{C! \Gamma(\alpha)} \left(\frac{\alpha \beta v}{\alpha + \alpha \beta v} \right)^C \left(\frac{\alpha}{\alpha + \alpha \beta v} \right)^\alpha \quad (10)$$

The joint distribution of tows and catches are shown in Table 11.

Table 10: Conditional probability distribution for catch (C) given the number of tows (q) with doubly truncated geometric sampling in a zero inflated negative binomial density field and tow volumes (v_i) vary between tows. For the case of two tows ($q = 2$), C_1 and C_2 correspond to the catches in the first and second tows, respectively, and $C_1 + C_2 = C$.

q	C	$\Pr(C q)$
2	$C = 1, 2, \dots$	$\frac{p(C_1; v_1)p(C_2; v_2)}{1-p(0; v_1)p(0; v_2)}$
$3 \leq q \leq q_{m-1}$	$C = 1, 2, \dots$	$\frac{p(C_q; v_q)}{1-p(0; v_q)}$
q_m	$C = 0$	$p(0; v_{q_m})(1 - \pi_0) + \pi_0$
	$C = 1, 2, \dots$	$p(C; v_{q_m})(1 - \pi_0)$

Table 11: Joint probability distribution for tows (q) and catch (C) with doubly truncated geometric sampling in a zero inflated negative binomial density field and tow volumes (v_i) vary between tows.

q, C	$\Pr(q, C)$
$q = 2, C = 1, 2, \dots$	$p(C_1; v_1) p(C_2; v_2) (1 - \pi_0)$
$3 \leq q \leq q_{m-1}, C = 1, 2, \dots$	$\prod_{i=1}^{q-1} p(0; v_i) p(C_q; v_q) (1 - \pi_0)$
$q = q_m, C = 0$	$\prod_{i=1}^{q_m} p(0; v_i) (1 - \pi_0) + \pi_0$
$q = q_m, C = 1, 2, \dots$	$\prod_{i=1}^{q_m-1} p(0; v_i) p(C_{q_m}; v_{q_m}) (1 - \pi_0)$

The log likelihood function (based on Table 11) is shown below.

$$\begin{aligned}
l(\pi_0, \delta) \propto & \sum_{i=1}^n I_{q_i=2} [\ln(p(C_1; v_1)) + \ln(p(C_2; v_2)) + \ln(1 - \pi_0)] + \\
& I_{3 \leq q_i \leq (q_m-1)} \left[\sum_{j=1}^{q_i} \ln(p(0; v_i)) + \ln(p(C_q; v_q)) + \ln(1 - \pi_0) \right] + \\
& I_{q=q_m, C_i=0} \ln \left[\prod_{i=1}^{q_m} p(0; v_i) (1 - \pi_0) + \pi_0 \right] + \\
& I_{q=q_m, C_i>0} \left[\sum_{j=1}^{q_m} \ln(p(0; v_i)) + \ln(p(C_{q_m}; v_{q_m})) + \ln(1 - \pi_0) \right]
\end{aligned} \tag{11}$$

R code for the negative log likelihood function is shown in Appendix C.3.

3.3 Simulation analysis of mles

A small factorial design simulation experiment was carried out to examine the performance of the maximum likelihood estimates of π_0 and $E[\delta]$, where $E[\delta] = \alpha\beta$, when catches were ZINB distributed. Four factors were considered: π_0 , δ , q_m , and n . The levels of each factor are shown in Table 3.3. Thus there are $4 \times 3 \times 2 \times 3 = 108$ “treatment” combinations. Tow volumes were randomly generated according to a Gamma(11.86, 261.29), which yields an average volume of 3100 m^3 and a standard deviation of 900. For each combination, 1000 samples were generated

Table 12: Factors and levels in simulation experiment to assess performance of estimates of π_0 and $\delta = \alpha\beta$ for ZINB catch data.

Factor	Levels
π_0	0.1, 0.3, 0.6, 0.9
δ	1e-3, 5e-4, 5e-5
q_m	5, 10
n	6, 9, 12

according to the negative binomial distribution, and maximum likelihood estimates of π_0 , and δ were calculated using the log likelihood function (11). The product $(1 - \pi_0)\delta$, namely the system-wide density, was also calculated using $(1 - \hat{\pi}_0)\hat{\delta}$.

Table 13 contains the average maximum likelihood estimate for π_0 and δ , and Table 14 has the corresponding standard deviations (thus estimates of the standard errors of the mles). As π_0 increases and δ decreases, the performance of the mles worsens. For example, when $\pi_0=0.9$, $\delta=5e-5$, $q_m=5$ and $n=6$ the average values of $\hat{\pi}_0$ was 0.78 and 1.8e-5 for $\hat{\delta}$ (see the next to last row in Table 13). The probability of all sampling locations failing to catch any fish was 72%. Figure 4 contains boxplots of the sampling distributions of the mles. Also shown are estimates based on, wrongly, assuming that the catches followed a zero inflated Poisson distribution; estimates of δ are considerably worse in that case. Increasing q_m and n does reduce the bias in estimates of δ somewhat, e.g., when $q_m=10$ and $n=12$, the probability of failing to catch any fish decreases to 38% and the average for $\hat{\delta}$ was 4.7e-5, however the average for $\hat{\pi}_0$ was 70% compared to the true value of 90%. Increasing the number of sampling locations had more effect on the standard errors than did increasing the maximum number of tows (Table 14). Results from a less extreme combination, are shown in Figure 5, where $\pi_0=0.3$, $\delta=5e-4$, $q_m=10$, but n still equals 6. In this case (see also Table 13) both the ZINB and ZIP estimates of π_0 and δ were relatively accurate.

Table 13: Average mles for π_0 and $E[\delta]$ based on 1000 simulations ZINB distributed catch data for each combination of π_0 , δ , q_m , and sample size n . $E[\delta]$ equaled $\alpha\beta$ where $\alpha = 1/CV^2$ and $\beta=CV^2\delta$ with $CV=0.4$.

π_0	δ	q_m	$\hat{\pi}_0$			$\hat{E}[\delta]$		
			$n=6$	$n=9$	$n=12$	$n=6$	$n=9$	$n=12$
0.1	0.001	5	0.1	0.11	0.1	0.001	0.001	0.001
0.1	0.001	10	0.1	0.1	0.1	0.001	0.001	0.001
0.1	5e-04	5	0.11	0.1	0.09	5e-04	5e-04	5e-04
0.1	5e-04	10	0.1	0.09	0.1	0.00051	5e-04	5e-04
0.1	5e-05	5	0.14	0.12	0.12	6.9e-05	6.4e-05	6.2e-05
0.1	5e-05	10	0.11	0.11	0.1	6.5e-05	6e-05	5.8e-05
0.3	0.001	5	0.3	0.31	0.31	0.00099	0.00099	0.001
0.3	0.001	10	0.3	0.3	0.31	0.001	0.001	0.001
0.3	5e-04	5	0.3	0.29	0.3	0.00051	5e-04	5e-04
0.3	5e-04	10	0.3	0.3	0.3	5e-04	5e-04	0.00051
0.3	5e-05	5	0.22	0.16	0.16	6.1e-05	5.7e-05	5.6e-05
0.3	5e-05	10	0.23	0.2	0.21	6.4e-05	5.9e-05	5.5e-05
0.6	0.001	5	0.6	0.59	0.6	0.00096	0.001	0.00099
0.6	0.001	10	0.61	0.6	0.61	0.00097	0.00099	0.001
0.6	5e-04	5	0.58	0.59	0.59	0.00047	5e-04	5e-04
0.6	5e-04	10	0.58	0.6	0.61	0.00048	5e-04	5e-04
0.6	5e-05	5	0.42	0.34	0.28	5.1e-05	5.6e-05	4.6e-05
0.6	5e-05	10	0.43	0.4	0.37	6.1e-05	6.2e-05	5.6e-05
0.9	0.001	5	0.9	0.9	0.9	0.00047	0.00064	0.00071
0.9	0.001	10	0.9	0.9	0.9	0.00047	6e-04	0.00073
0.9	5e-04	5	0.88	0.88	0.88	0.00023	0.00031	0.00035
0.9	5e-04	10	0.9	0.9	0.9	0.00024	0.00031	0.00037
0.9	5e-05	5	0.78	0.74	0.7	1.8e-05	2.5e-05	2.9e-05
0.9	5e-05	10	0.8	0.74	0.7	3.2e-05	4.3e-05	4.7e-05

Table 14: Standard errors of mles for π_0 and $E[\delta]$ (see caption for Table 13).

π_0	δ	q_m	$\hat{\pi}_0$			$\hat{E}[\delta]$		
			$n=6$	$n=9$	$n=12$	$n=6$	$n=9$	$n=12$
0.1	0.001	5	0.12	0.11	0.09	0.00021	0.00017	0.00015
0.1	0.001	10	0.12	0.1	0.09	0.00022	0.00017	0.00015
0.1	5e-04	5	0.13	0.1	0.08	0.00013	0.00011	9.4e-05
0.1	5e-04	10	0.12	0.1	0.09	0.00013	0.00011	9.2e-05
0.1	5e-05	5	0.26	0.23	0.2	5.9e-05	4.9e-05	3.8e-05
0.1	5e-05	10	0.18	0.17	0.15	4.2e-05	3.2e-05	2.7e-05
0.3	0.001	5	0.18	0.15	0.13	0.00026	2e-04	0.00018
0.3	0.001	10	0.18	0.15	0.13	0.00025	0.00021	0.00017
0.3	5e-04	5	0.19	0.16	0.13	0.00017	0.00013	0.00011
0.3	5e-04	10	0.19	0.15	0.13	0.00016	0.00013	0.00011
0.3	5e-05	5	0.33	0.27	0.26	6.1e-05	5.2e-05	4.9e-05
0.3	5e-05	10	0.27	0.23	0.22	5.3e-05	4.1e-05	3.8e-05
0.6	0.001	5	0.21	0.16	0.14	0.00042	0.00029	0.00024
0.6	0.001	10	0.2	0.17	0.14	0.00041	0.00031	0.00025
0.6	5e-04	5	0.21	0.16	0.14	0.00025	0.00019	0.00016
0.6	5e-04	10	0.2	0.17	0.14	0.00023	0.00019	0.00015
0.6	5e-05	5	0.43	0.4	0.37	7.3e-05	7.8e-05	6.1e-05
0.6	5e-05	10	0.38	0.35	0.33	7e-05	6.8e-05	5.4e-05
0.9	0.001	5	0.13	0.1	0.1	6e-04	0.00061	0.00057
0.9	0.001	10	0.12	0.1	0.09	6e-04	6e-04	0.00058
0.9	5e-04	5	0.18	0.16	0.15	0.00033	0.00034	0.00034
0.9	5e-04	10	0.12	0.1	0.09	0.00033	0.00034	0.00032
0.9	5e-05	5	0.38	0.41	0.43	5.3e-05	6.6e-05	6.8e-05
0.9	5e-05	10	0.34	0.38	0.4	6.9e-05	7.4e-05	7e-05

Figure 4: ZINB-varying volumes. Boxplots of sampling distributions of $\hat{\pi}_0$, $\hat{\delta}$, and $(1 - \hat{\pi}_0)\hat{\delta}$ under the ZINB process model (based on 1000 simulations) when $\pi_0=0.9$, $\delta=5e-5$, $q_m=5$, and $n=6$. The blue lines mark the true values and the red dashed lines are the average of the mles. The dotted purple lines mark the means using the ZIP estimator (wrongly assuming a ZIP distribution).

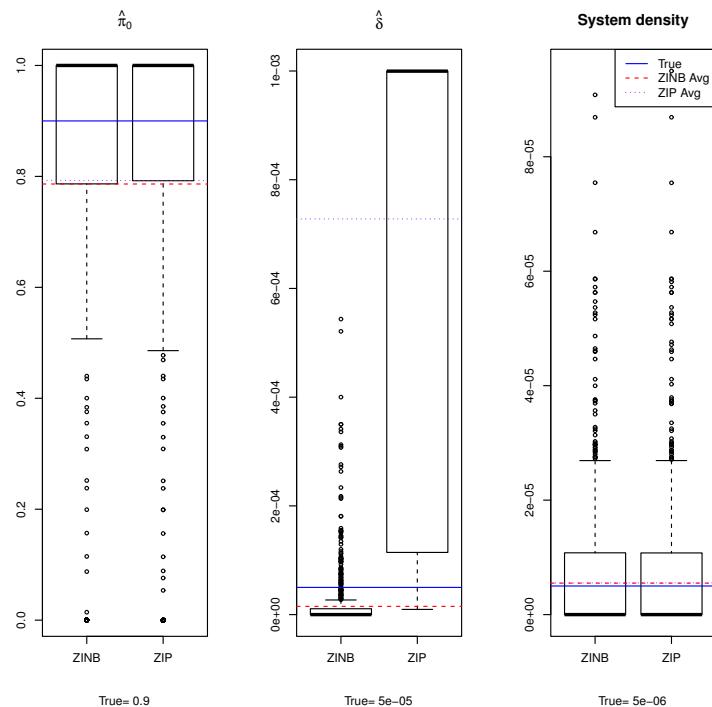
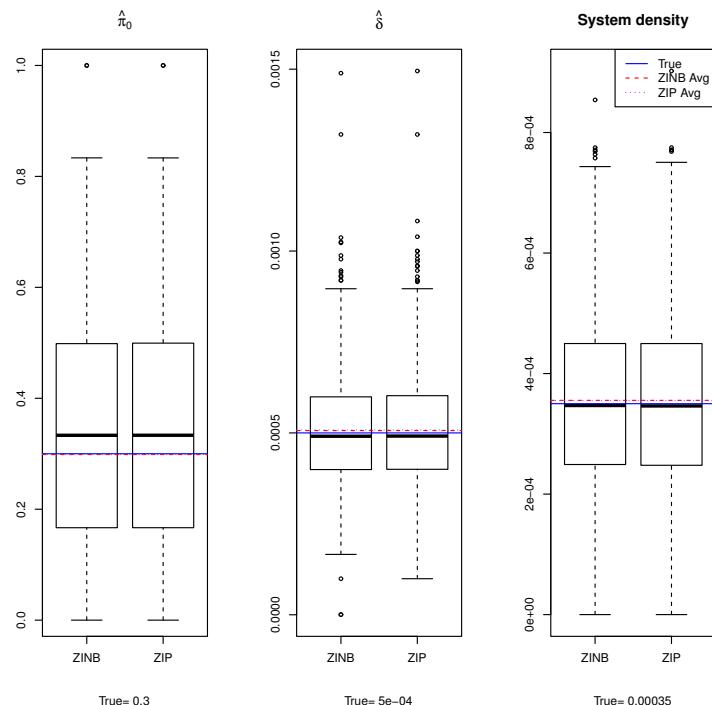


Figure 5: ZINB-varying volumes. Boxplots of sampling distributions of $\hat{\pi}_0$, $\hat{\delta}$, and $(1 - \hat{\pi}_0)\hat{\delta}$ under the ZINB process model (based on 1000 simulations) when $\pi_0=0.3$, $\delta=5e-4$, $q_m=10$, and $n=6$. The blue lines mark the true values and the red dashed lines are the average of the mles. The dotted purple lines mark the means using the ZIP estimator (wrongly assuming a ZIP distribution).



A Proof of equivalency of Poisson-Gamma hierarchical catch model and negative binomial catch model

Starting with the same hierarchical model as in Section 3.1,

$$\begin{aligned}\alpha &= \frac{1}{CV^2} \\ \beta &= CV^2\mu_\delta \\ \delta_i &\sim \text{Gamma}(\alpha, \beta), \quad i = 1, \dots, n \\ C_{ij} &\sim \text{Poisson}(\delta_i v), \quad j = 1, \dots, q_i\end{aligned}$$

the distribution for catch given α and β is equivalent to a negative binomial distribution¹:

$$C_{ij} \sim \text{NegBinom}(\alpha\beta v, \alpha).$$

A proof of this equivalency follows:

$$P(C|\alpha, \beta) = \int_0^\infty P(C, \delta|\alpha, \beta) d\delta = \int_0^\infty P(C|\delta) P(\delta|\alpha, \beta) d\delta \quad (12)$$

$$= \int_0^\infty \frac{\exp(-\delta v)(\delta v)^C}{C!} \frac{\delta^{\alpha-1} \exp(-\delta/\beta)}{\beta^\alpha \Gamma(\alpha)} d\delta \quad (13)$$

$$= \frac{v^C}{C! \beta^\alpha \Gamma(\alpha)} \int_0^\infty \exp(-\delta(v + 1/\beta)) \delta^{C+\alpha-1} d\delta \quad (14)$$

$$= \frac{v^C}{C! \beta^\alpha \Gamma(\alpha)} \left(\frac{\beta}{v\beta+1} \right)^{C+\alpha} \Gamma(C+\alpha) \int_0^\infty \frac{\exp\left(-\frac{\delta}{\left(\frac{\beta}{v\beta+1}\right)}\right) \delta^{C+\alpha-1}}{\left(\frac{\beta}{v\beta+1}\right)^{C+\alpha} \Gamma(C+\alpha)} d\delta \quad (15)$$

$$= \frac{v^C}{C! \beta^\alpha \Gamma(\alpha)} \left(\frac{\beta}{v\beta+1} \right)^{C+\alpha} \Gamma(C+\alpha) \quad (16)$$

$$= \frac{\Gamma(C+\alpha)}{C! \Gamma(\alpha)} \left(\frac{\alpha\beta v}{\alpha + \alpha\beta v} \right)^C \left(\frac{\alpha}{\alpha + \alpha\beta v} \right)^\alpha. \quad (17)$$

B R code for the distribution of q

```
#---- Theoretical results for doubly truncated geometric
ZI.doubly.trunc.geom.pmf <- function(pi.0, theta, qmax) {
  out <- numeric(qmax)
  out[2] <- (1-theta^2)*(1-pi.0)
  for(i in 3:(qmax-1)) {
    out[i] <- theta^(i-1)*(1-theta)*(1-pi.0)
  }
}
```

¹A negative binomial(μ, θ) distribution is used where the parameterization is in terms of the mean, μ , and the dispersion parameter θ , such that $E[X] = \mu$, $Var[X] = \mu + \mu^2/\theta$. The pdf in this case is the following:

$$\frac{\Gamma(x+\theta)}{x!\Gamma(\theta)} \left(\frac{\mu}{\theta+\mu} \right)^x \left(\frac{\theta}{\theta+\mu} \right)^\theta$$

```

out[qmax] <- theta^(qmax-1)*(1-pi.0)+pi.0
out <- out[-1]
names(out) <- 2:qmax
return(out)
}

# Expected number of tows
ZI.expected.tows.double.geom <- function(pi.0,v.bar,delta,qmax) {
  exp.non.ZI <- expected.tows.double.geom(v.bar=v.bar,delta=delta,qmax=qmax)
  out <- (1-pi.0)*exp.non.ZI + pi.0*qmax
  return(out)
}

# Variance of number of tows
ZI.var.tows.double.geom <- function(pi.0,v.bar,delta,qmax) {
  exp.non.ZI <- expected.tows.double.geom(v.bar=v.bar,delta=delta,qmax=qmax)
  exp.sq.non.ZI <- exp.tows.sq.double.geom(v.bar=v.bar,delta=delta,qmax=qmax)
  p1 <- (1-pi.0)*exp.sq.non.ZI+qmax^2*pi.0
  p2 <- exp.non.ZI^2*(1-pi.0)^2 + 2*exp.non.ZI*(1-pi.0)*qmax*pi.0+qmax^2*pi.0^2
  out <- p1-p2
  return(out)
}

#--- Empirically verify accuracy of E[q] and Var[q] calculations
delta <- 0.5/10000
v.bar <- 7000
theta <- exp(-v.bar*delta)
qmax <- 10
pi.0 <- 0.8

num.sims <- 10000
out <- numeric(num.sims)
verbose <- FALSE
for(i in 1:num.sims) {
  zero.d <- rbinom(n=1,size=1,prob=pi.0)
  if(zero.d==1) {
    out[i] <- qmax
  } else { # not zero density

    notok <- TRUE
    q <- 2
    # see if succeed in 1st two tows
    if(runif(n=1,min=0,max=1) <= (1-theta^2)) {
      notok <- FALSE
    }
    while(notok & q<qmax) {
      q <- q + 1
      if(runif(n=1,min=0,max=1) > theta) {
        notok <- FALSE
      }
    }
    if(verbose) {
      if(notok==FALSE) {
        cat("iter=",i,"success, q=",q,"\n")
      } else {
        cat("iter=",i,"fail, q=",q,"\n")
      }
    }
  }
}

```

```

    out[i] <- q
}
} #---- end of loop

cat("Empirical mean=",mean(out),
  "Theoretical=",ZI.expected.tows.double.geom(pi.0=pi.0,v.bar=v.bar,delta=delta,
  qmax=qmax),"\n")

# Empirical mean= 8.6924 Theoretical= 8.715861

cat("Empirical std dev",sd(out),"Theoretical std=",
  sqrt(ZI.var.tows.double.geom(pi.0=pi.0,v.bar=v.bar,delta=delta,qmax=qmax)),"\n")

# Empirical std dev 2.769215 Theoretical sd= 2.755594

cat("Comparison of pmf: empirical vs theoretical \n")
temp <- rbind(table(out)/sum(table(out)),
  ZI.doubly.trunc.geom.pmf(pi.0=pi.0,theta=exp(-delta*v.bar),qmax=qmax))
print(temp)
#          2         3         4         5         6         7         8         9        10
#[1,] 0.0994000 0.03310000 0.02120000 0.0143000 0.0112000 0.007800000 0.005100000 0.003600000 0.8043000
#[2,] 0.1006829 0.02932951 0.02066816 0.0145646 0.0102635 0.007232568 0.005096705 0.003591587 0.8085704

hist(out,main="Dist'n of q",xlab="q")
legend("center",legend=c(paste("Emp.mean=",round(mean(out),1),
  "Theory=",round(ZI.expected.tows.double.geom(pi.0=pi.0,v.bar=v.bar,
  delta=delta,qmax=qmax),1)),
  paste("Emp.sd=",round(sd(out),1),"Theory=",
  round(sqrt(ZI.var.tows.double.geom(pi.0=pi.0,v.bar=v.bar,delta=delta,
  qmax=qmax)),1))))
```

C R code for calculating mle of (π_0, δ) .

C.1 Constant volume

The following code assumes a constant volume is sampled with each tow.

```

#--- Case 1: ZIP constant volume -----
#   Negative Log Likelihood function for pi.0 and delta -----
logit <- function(x) log(x/(1-x))
expit <- function(x) exp(x)/(1+exp(x))

neg.logL.ZIP.double.geom.const.v <- function(theta,v.bar,qmax,q.vec,catch.vec) {
  # corresponds to Eq'n 4 in TN 23
  pi.0 <- expit(theta[1])
  delta <- exp(theta[2])

  vd     <- v.bar*delta
  n      <- length(q.vec)
  q2.cases       <- q.vec==2
  qmid.cases    <- (q.vec >= 3) & (q.vec <= (qmax-1))
  qmax.zeroC.cases <- (q.vec==qmax) & (catch.vec==0)
  qmax.posC.cases <- (q.vec==qmax) & (catch.vec>0)
  p1 <- p2 <- p3 <- p4 <- 0
  if(sum(q2.cases)>0) {
```

```

p1 <- sum((-2*vd+catch.vec*log(2*vd)+log(1-pi.0))[q2.cases])
}
if(sum(qmid.cases)>0) {
  p2 <- sum((-q.vec*vd+catch.vec*log(vd)+log(1-pi.0))[qmid.cases])
}
if(sum(qmax.zeroC.cases)>0) {
  p3 <- log(exp(-qmax*vd)*(1-pi.0)+pi.0)*sum(qmax.zeroC.cases)
}
if(sum(qmax.posC.cases)>0) {
  p4 <- sum((-qmax*vd+catch.vec*log(vd)+log(1-pi.0))[qmax.posC.cases])
}
out <- -(p1+p2+p3+p4)
#gr <- logL.ZI.gradient(theta=theta,v.bar=v.bar,qmax=qmax,
#  q.vec=q.vec,catch.vec=catch.vec,verbose=FALSE)
#attr(out, "gradient") <- gr
return(out)
}

#----- Simulating tows and catches -----
pi.0      <- 0.3
v.bar      <- 7000
delta      <- 5/10000
num.samples <- 5
catch.vec <- q.vec <- delta.vec <- numeric(num.samples)

for(k in 1:num.samples) {
  #-- simulate zero density
  zero.d <- rbinom(n=1,size=1,prob=pi.0)
  if(zero.d==1) {
    catch.vec[k] <- 0
    q.vec[k]      <- qmax
  } else { # not zero density
    no.catch <- TRUE
    q.vec[k] <- qmax
    catch.vec[k] <- rpois(n=1,lambda=2*v.bar*delta)
    if(catch.vec[k]>0) {
      q.vec[k] <- 2
    } else {
      q.vec[k] <- 2
      while(no.catch & q.vec[k] <= (qmax-1)) {
        q.vec[k]      <- q.vec[k] + 1
        catch.vec[k] <- rpois(n=1,lambda=v.bar*delta)
        if(catch.vec[k] > 0) {
          no.catch <- FALSE
        }
      }
    }
  }
}

#--- mle computation
out <- nlm(f=neg.logL.ZI.double.geom,p=c(logit(0.5),log(0.001)),
hessian=TRUE,
v.bar=v.bar,qmax=qmax,q.vec=q.vec,catch.vec=catch.vec,verbose=FALSE)
pi.0.hat   <- expit(out$estimate[1])
delta.hat   <- exp(out$estimate[2])

```

```

##--calculate std errors
fisher_info    <- solve(out$hessian)
var.mle        <- diag(fisher_info)

var.logit.pi.0 <- var.mle[1]
se.pi.0        <- sqrt(exp(out$estimate[1])^2/((1+exp(out$estimate[1]))^4)*
                     var.logit.pi.0)
var.log.delta   <- var.mle[2]
se.delta       <- sqrt(exp(out$estimate[2])^2*var.log.delta)

##--compare mles to true values
# pi.0= 0.3   Est= 0.163   SE= 0.205
# delta= 5e-05 Est= 4.44e-05 SE= 3.06e-05

##--contour plot of negative log likelihood with estimate and true value
pi.set      <- seq(0.1,0.6,length=20)
delta.set   <- seq(0.4,1,length=20)/10000
out.logL <- matrix(data=NA,nrow=length(pi.set),ncol=length(delta.set))
for(i in 1:length(pi.set)) {
  for(j in 1:length(delta.set)) {
    out.logL[i,j] <- neg.logL.ZI.double.geom(theta=c(logit(pi.set[i]),
      log(delta.set[j])),
      v.bar=v.bar,qmax=qmax,
      q.vec=q.vec,catch.vec=catch.vec,verbose=FALSE)
  }
}

contour(x=pi.set,y=delta.set,z=out.logL,
  xlab="pi.0",ylab="delta",main="")
points(pi.0,delta,cex=1,col="blue",pch="T")
points(pi.0.hat,delta.hat,cex=1,col="red",pch="E")
segments(pi.0.hat-2*se.pi.0,delta.hat,pi.0.hat+2*se.pi.0,delta.hat,col="red")
segments(pi.0.hat,delta.hat-2*se.delta,pi.0.hat,delta.hat+2*se.delta,col="red")

```

C.2 ZIP-Varying volume

The following code allows the volume sampled to vary between tows.

```

##### Negative Log Likelihood function for pi.0 and delta ----

neg.logL.ZI.double.geom.varying.v <- function(theta,v.matrix,qmax,catch.vec,
  verbose=FALSE) {
  #v.matrix is rectangular n X qmax matrix, with NAs to indicate where
  # the number of tows ended
  pi.0 <- expit(theta[1])
  delta <- exp(theta[2])
  n     <- length(catch.vec)

  #Tally the total volume sampled at each location
  v.totals <- apply(v.matrix,1,sum,na.rm=TRUE)
  vttl.d    <- v.totals*delta

  #Extract the volume for the last tow
  v.extend <- cbind(v.matrix,NA)
  temp      <- is.na(v.extend)

```

```

row.match <- function(x) {
  match(TRUE,x)
}
last.tow <- apply(temp,1,row.match)-1
last.vol <- v.matrix[cbind(1:n,last.tow)]

if(verbose) {
  cat("pi.0=",pi.0,"delta=",delta,"\n")
  cat("total volume= \n");print(v.totals)
  cat("last volume= \n");print(last.vol)
}

#Indicator variables for q2, 3 to q_{m-1}, q_m (w/ or w/o catch)
q2.cases      <- last.tow==2
qmid.cases     <- (last.tow >= 3) & (last.tow <= (qmax-1))
qmax.posC.cases <- (last.tow==qmax) & (catch.vec>0)
qmax.zeroC.cases <- (last.tow==qmax) & (catch.vec==0)

if(verbose) {
  cat("q2=",sum(q2.cases),"qmid=",sum(qmid.cases),
  "qmax.posC=",sum(qmax.posC.cases),"qmax.zeroC=",sum(qmax.zeroC.cases),"\n")
}

p1 <- p2 <- p3 <- p4 <- 0
# catch in first two tows
if(sum(q2.cases)>0) {
  p1 <- sum((-vttl.d+catch.vec*log(vttl.d)+log(1-pi.0))[q2.cases])
}
# catch in tows 3 to qmax-1
if(sum(qmid.cases)>0) {
  p2 <- sum((-vttl.d+catch.vec*log(last.vol*delta)+log(1-pi.0))[qmid.cases])
}
# cases with max tows and positive catch
if(sum(qmax.posC.cases)>0) {
  p3 <- sum((-vttl.d+catch.vec*log(last.vol*delta)+log(1-pi.0))[qmax.posC.cases])
}
# cases with max tows and zero catch
if(sum(qmax.zeroC.cases)>0) {
  p4 <- sum(log(exp(-vttl.d)*(1-pi.0)+pi.0)[qmax.zeroC.cases])
}

if(verbose) {
  cat("p1=",p1,"p2=",p2,"p3=",p3,"p4=",p4,"\\n")
}

out <- -(p1+p2+p3+p4)
if(verbose) cat("out=",out,"\\n")

#gr <- logL.ZI.gradient(theta=theta,v.bar=v.bar,qmax=qmax,
#  q.vec=q.vec,catch.vec=catch.vec,verbose=FALSE)
#attr(out, "gradient") <- gr
return(out)
}

```

C.3 ZINB-Varying volume

The following code allows the volume sampled to vary between tows with catches following a zero inflated negative binomial distribution.

```

---- Case 3: overdispersed catch (ZINB) and varying volume ----
# Negative Log Likelihood function for pi.0, alpha, beta (ZINB):
neg.logL.ZINB.double.geom.varying.v <- function(theta,volume.matrix,qmax,
  catch.matrix,verbose=FALSE) {
  #volume.matrix is rectangular n X qmax matrix, with NAs to indicate where
  # the number of tows ended
  pi.0 <- expit(theta[1])
  alpha <- exp(theta[2])
  beta <- exp(theta[3])
  mu.delta <- alpha*beta
  n       <- nrow(catch.matrix)

  #Tally the total volume sampled at each location
  v.totals <- apply(volume.matrix,1,sum,na.rm=TRUE)
  vttl.d     <- v.totals*mu.delta

  #Extract the volumes for the last tow
  v.extend <- cbind(volume.matrix,NA)
  temp      <- is.na(v.extend)
  row.match <- function(x) {
    match(TRUE,x)
  }
  last.tow <- apply(temp,1,row.match)-1
  last.vol <- volume.matrix[cbind(1:n,last.tow)]
  first.vol <- volume.matrix[,1]

  #Extract the catches for the first and last tows
  first.catch <- catch.matrix[,1]
  last.catch <- catch.matrix[,last.tow]

  #Calculate matrix of NegBin probabilities, per tow
  negbin.probs.matrix <- dnbinom(x=catch.matrix,size=alpha,
    mu=alpha*beta*volume.matrix,log=FALSE)
  negbin.probs.log.matrix <- log(negbin.probs.matrix)

  if(verbose) {
    cat("pi.0=",pi.0,"alpha=",alpha,"beta=",beta,"E[delta]=",mu.delta,"\n")
  }

  #Indicator variables for q=2, 3 to q_{m-1}, q_m (w/ or w/o catch)
  q2.cases      <- last.tow==2
  qmid.cases    <- (last.tow >= 3) & (last.tow <= (qmax-1))
  qmax.zeroC.cases <- (last.tow==qmax) & (catch.matrix[,qmax]==0)
  qmax.posC.cases <- (last.tow==qmax) & (catch.matrix[,qmax]>0)

  if(verbose) {
    cat("q2=",sum(q2.cases),"qmid=",sum(qmid.cases),
    "qmax.zeroC=",sum(qmax.zeroC.cases),"qmax.posC=",sum(qmax.posC.cases),"\\n")
  }

  p1 <- p2 <- p3 <- p4 <- 0
  # catch in first two tows: need to know catch in each tow
  if(sum(q2.cases)>0) {
    p1 <- sum((negbin.probs.log.matrix[,1] +
      negbin.probs.log.matrix[,2] + log(1-pi.0))[q2.cases])
  }
  # catch in tows 3 to qmax-1
  if(sum(qmid.cases)>0) {
    p2 <- sum((apply(negbin.probs.log.matrix,1,sum,na.rm=TRUE) +
      log(1-pi.0))[qmid.cases])
  }
}

```

```
}

# cases with max tows and zero catch
if(sum(qmax.zeroC.cases)>0) {
  p3 <-  sum(log(apply(negbin.probs.matrix,1,prod,na.rm=TRUE)*(1-pi.0)+
    pi.0)[qmax.zeroC.cases])
}

# cases with max tows and positive catch
if(sum(qmax.posC.cases)>0) {
  p4 <-  sum((apply(negbin.probs.log.matrix,1,sum,na.rm=TRUE)+
    log(1-pi.0))[qmax.posC.cases])
}

out <- -(p1+p2+p3+p4)
if(verbose)  cat("p1=",p1,"p2=",p2,"p3=",p3,"p4=",p4,"out=",out,"\n")
return(out)
}
```